# On parametrizing the rain drop size distribution

Ziad S. Haddad
Stephen L. Durden
Eastwood Im

Jet Propulsion Laboratory, California Institute of Technology
Pasadena, California

December 1, 1994

## Abstract

This paper addresses the problem of finding a parametric form for the rain drop size distribution which 1) is an appropriate model for tropical rainfall, and 2) involves statistically independent parameters. Such a parametrization is derived in this paper. Two of the resulting three "canonical" parameters turn out to vary relatively little, thus making the parametrization particularly useful for remote sensing applications.

# 1 introduction

Since Marshall and Palmer's pioneering 1948 paper, much attention has been focused on obtaining relatively simple analytic expressions involving as small a number of parameters as possible to model measured drop size distributions (DSD's). The three-parameter Γ-distribution model

$$N(D) = N_0 D^{\mu} e^{-\Lambda D} \tag{1}$$

proposed by Ulbrich (1983) has been tested using different data sets (see e.g. Kozu and Nakamura, 1991, Goddard and Cherry, 1984, and Ulbrich, 1983), and it has proved to be sufficiently versatile to fit most data satisfactorily, as long as one is willing to allow a relatively wide range of values for the parameter $\mu$.

However, Ulbrich (1983) pointed out, and the present work confirms, that the parameters $N_0$, $\mu$ and $\Lambda$ are not mutually independent. In practice, this makes the representation (1) difficult to use in rain retrieval algorithms. To illustrate the problem, suppose one has measurements of a rain-related quantity $Z = Z(a)$ at various altitudes $a$ in the atmosphere. One may then try to determine the distribution $N(D; a)$ at the corresponding altitudes. A priori, all three parameters in (1) may vary with $a$. Yet given one's single observed quantity $Z$, it is unrealistic to expect to successfully determine, at each altitude, the triple $(N_0, \mu, \Lambda)$ that produced the observed value of $Z$. In this case, one way to circumvent this problem is to assume that the typically-less-variable parameters are constant, e.g. make $N_0$ and $\mu$ constant, and determine $\Lambda(a)$ as a function of the observed $Z(a)$. The specific (constant) values of $N_0$ and $\mu$ need not be known beforehand; one may try to determine them using ancillary observations or archived historical data. The problem with this approach is that it makes little sense to assume $N_0$ constant and let $\Lambda$ change according to the observation, when one already knows that $N_0$ and $\Lambda$ are strongly correlated.

It is therefore very useful to derive an expression like (1) but involving statistically independent parameters, ones that are preferably physically meaningful. That is the aim of this paper.

# 2 Statistical analysis of the Darwin data

The data analyzed were measured by a Joss-Waldvogel disdrometer (Joss and Waldvogel, 1967, Sheppard and Joe, 1994) located at Berrimah near Darwin, Australia. The measurements were taken during the southern-hemisphere summer seasons of 1988 1989 and

1989–1990. The disdrometer recorded the number of drops in each of 20 drop-diameter bins as in table 1, reporting a sample distribution every 30 seconds. In order to fit a model

| Bin number $i$ | Actual drop diameters | Reported diameter $D_i$ |
|---|---|---|
| 1 | 0.36 - 0.48 mm | 0.42 mm |
| 2 | 0.48 - 0.6 mm | 0.54 mm |
| 3 | 0.6 - 0.72 mm | 0.66 mm |
| 4 | 0.72 - 0.84 mm | 0.78 mm |
| 5 | 0.84 - 0.96 mm | 0.9 mm |
| 6 | 0.96 - 1.2 mm | 1.08 mm |
| 7 | 1.2 - 1.44 mm | 1.32 mm |
| 8 | 1.44 - 1.68 mm | 1.56 mm |
| 9 | 1.68 - 1.92 mm | 1.8 mm |
| 10 | 1.92 - 2.16 mm | 2.04 mm |
| 11 | 2.16 - 2.52 mm | 2.34 mm |
| 12 | 2.52 - 2.88 mm | 2.7 mm |
| 13 | 2.88 - 3.24 mm | 3.06 mm |
| 14 | 3.24 - 3.6 mm | 3.42 mm |
| 15 | 3.6 - 3.96 mm | 3.78 mm |
| 16 | 3.96 - 4.44 mm | 4.2 mm |
| 17 | 4.44 - 4.92 mm | 4.68 mm |
| 18 | 4.92 - 5.4 mm | 5.16 mm |
| 19 | 5.4 - 6.0 mm | 5.7 mm |
| 20 | 6.0 - 6.6 mm | 6.3 mm |

Table 1: Disdrometer bin values (the 20% discrepancy with the values in Joss and Waldvogel, 1967, is due to a mis-calibration of the instrument at Darwin).

such as (1) to any such sample distribution, one could proceed in several ways. One way is to express the predicted moments of the DSD as functions of $(N_0, \mu, \Lambda)$, then use three suitable sample moments computed from one's observations to perform the inversion and deduce the values of the three parameters. This approach is quite unappealing because its estimates would depend on the moments used, and, more important, because sample moments (and, a fortiori, complicated functions of several of them) are biased estimates of the actual moments. A simpler way is to find those values of $(N_0, \mu, \Lambda)$ which minimize the sum of the squared differences between the observed counts and those predicted by (1). This least-squares approach implicitly assumes that the difference between the observation and one's model is entirely due to white noise evenly spread among the sampling bins. While such an assumption is appealingly simple, it does not allow one to use all the information at hand. A maximum-likelihood approach does. Indeed, one can view the drop-size distribution $N(D)$ as the product of a drop-size *density function* $\mathcal{P}_{\mu,\Lambda}(D) = \Lambda^{\mu+1} D^{\mu} e^{-\Lambda D}/\Gamma(\mu+1)$, which depends on $\mu$ and $\Lambda$ only, with the total number of drops $N_0 \Gamma(\mu+1)\Lambda^{-\mu-1}$. Since the latter is directly related to the observed total count, the problem of estimating $\mu$ and $\Lambda$ reduces to finding the values of these two parameters that maximize the likelihood

$$\prod_{j=1}^{20} \left(\mathcal{P}_{\mu,\Lambda}(D_j)\right)^{N_j} \tag{2}$$

of obtaining the counts $N_j$ that were computed from the observations (using equation 3 in Sheppard and Joe, 1994), with $D_j$ as in table 1. Instead of the abstract parameters $N_0$, $\mu$ and $\Lambda$, we used the more physically meaningful variables

$$D^* = \text{mass-weighted mean drop diameter} = \frac{\mu+4}{\Lambda}\ \text{mm} \tag{3}$$

$$s^* = \text{relative mass-weighted r.m.s. deviation of drop diameter} = \frac{1}{\sqrt{\mu+4}} \tag{4}$$

$$R = \text{instantaneous rain rate} = 7.1 \times 10^{-3}\, \frac{\Gamma(\mu+4.67)}{\Lambda^{\mu+4.67}}\, N_0\ \text{mm/hr}, \tag{5}$$

Relation (5) was obtained assuming that $D$ is in millimeters and that the fall velocity $v$ of a drop of diameter $D$ mm is $v = 3.78 D^{0.67}$ m/sec (Atlas and Ulbrich 1977). The maximum-likelihood estimates are shown in the pairwise-scatter diagrams of figures 1a–1f. Since the measurements made during very light rain are unreliable because of the small sample size, we imposed a lower-bound condition on $R$. The particular value of 0.7 mm/hr was chosen because it corresponds to the projected Tropical Rainfall Measuring Mission radar's sensitivity (Kawanishi et al, 1993). The values of the various conditional correlation coefficients

| 88-89 \ 89-90 | $R$ | $D^*$ | $s^*$ |
|---|---|---|---|
| $R$ | | 0.63 | -0.62 |
| $D^*$ | 0.57 | | 0.08 |
| $s^*$ | -0.45 | 0.1 | |

Table 2: Correlation coefficients for the 88-89 and 89-90 seasons.

(conditioned on $R > 0.7$ mm/hr) are given in table 2. As one might have expected, there is no significant correlation between the mean drop diameters $D^*$ and the relative mean variance $s^*$ of the diameters. However, the data shows that both these quantities are rather strongly correlated with the rain rate. As to Ulbrich's original variables, the most striking correlation is that of $N_0$ with A: their correlation coefficient is 0.92 for the 1989 89 season, and 0.93 for 1989 90.

A popular way to make mathematically explicit the interdependences which underlie the observed correlations is to use power-law regressions and express one variable in terms of another, e.g. try to find intervals for $a$, $b$, $c$ and $d$ such that $N_0 = a\mu^b$ or $A = cR^d$ (see e.g. Ulbrich, 1983 and 1992). The problems with such an approach are that one then artificially introduces new coefficients (namely $a$, $b$, $c$ and $d$) which are not related to any of the original variables in a unique way, and whose mutual covariances are therefore impossible to determine. Since, in addition, such power-laws produce far more unknowns than one started with, a more efficient and consistent approach such as a simple (judicious) change of variables should prove more useful.

The simplest way to change variables so as to end up with an independent set is to find the (orthogonal) eigenvectors of the covariance matrix. However, the variables produced using such an approach will not be physically meaningful. In addition, it is very desirable to specifically retain the rain rate $R$ as one of the 3 variables since it is one of the quantities of most interest. So rather than diagonalize the covariance matrix, let us decide to keep $R$ as the first variable, and successively modify $D^*$ then $s^*$ in order to end up with a set of uncorrelated and, one hopes, jointly (log-)normal variables. The scatter diagrams 1c and 1d suggest that a linear change of variables

$$\log(D^*) = \log(D') + \alpha \log(R) \tag{6}$$

might allow one to replace $D^*$ with a new variable $D'$ such that $D'$ and $R$ are uncorrelated (and nearly jointly log-normal), if the slope $\alpha$ is chosen to make the correlation 0, i.e. to satisfy

$$\mathcal{E}\{\log(D^*)\log(R)\} - \alpha\,\mathcal{E}\{\log(R)^2\} = \mathcal{E}\{\log(D^*)\}\mathcal{E}\{\log(R)\} - \alpha\,\mathcal{E}\{\log(R)\}^2 \tag{7}$$

Using the 1988 89 data to estimate the second-order moments in (7), one finds that $\alpha$ should equal 0.1336, while the 1989 90 sample moments give the value 0.128. Retaining two significant digits gives a consistent value

$$\alpha = 0.13 \tag{8}$$

As to $s^*$, the scatter diagrams 1c and 1f suggest that the $s^*$-$R$ correlation is due mostly to data corresponding to high rain rates. In fact, when conditioned on $R < 9$ mm/hr, the $s^*$ $R$ conditional correlation coefficient drops from the original -0.45 to 0.0024 for the 88 89 season, and from -0.62 to -0.16 for the 89 90 data. Therefore, rather than a linear change of variables, a quadratic form

$$\log(s^*) = \log(s'') + \beta \log(R)^2 \tag{9}$$

seems more suitable. The value of $\beta$ that will make the correlation between $R$ and the new variable $s''$ zero can be derived as before. In this case, one finds

$$\beta = -0.02 \tag{10}$$

Finally, one needs to replace $s''$ by a variable which is uncorrelated with either $R$ or $D'$. Calling the new variable $s'$, if it is defined by a linear change of variables

$$\log(s'') = \log(s') + \gamma \log(D') \tag{11}$$

it will be automatically uncorrelated with $R$ (because $s''$ and $D'$ are), so it suffices to choose $\gamma$ in such a way that $s'$ and $D'$ are uncorrelated. Proceeding as before, one finds

$$\gamma = 0.35 \tag{12}$$

Thus our three uncorrelated variables are

$$R \;=\; \text{instantaneous rain rate} \tag{13}$$

$$D' \;=\; D^* R^{-0.13} \tag{14}$$

$$\text{and } s' \;=\; s^* D'^{-0.35} R^{0.02\log(R)} \tag{15}$$

The scatter diagrams in figures 2a-2f show the values of these new variables for the Darwin DSD's. The marginal statistics for each individual variable are summarized in table 3. It

| | Mean | | Standard Deviation | |
|---|---|---|---|---|
| | 1988-89 | 1989-90 | 1988-89 | 1989-90 |
| $D'$ | 1.51 | 1.56 | 0.33 | 0.36 |
| $s'$ | 0.337 | 0.355 | 0.037 | 0.038 |
| $R$ | 8.2 | 15.57 | 15.57 | 25.18 |
| $\log(D')$ | 0.387 | 0.42 | 0.22 | 0.22 |
| $\log(s')$ | -0.98 | -1.04 | 0.012 | 0.014 |
| $\log(R)$ | 1.31 | 1.14 | 1.13 | 1.38 |

Table 3: Marginal statistics of $R$, $D'$ and $s'$.

is quite encouraging to note that the standard deviation of $s'$ seems very small. Even the variance of $D'$ is relatively small, in spite of the large variability evident in the rain rate itself. Finally, table 4 confirms that the correlation coefficients are all negligibly small. For jointly log-normal variables, the vanishing of the correlation coefficients is equivalent to the mutual independence of the variables themselves. Thus one can conclude that the Darwin

| 88-89 \ 89-90 | $R$ | $D'$ | $s'$ |
|---|---|---|---|
| $R$ | | -0.15 | 0.058 |
| $D'$ | 0.044 | | 0.18 |
| $s'$ | 0.075 | -0.16 | |

Table 4: Correlation coefficients of $R$, $D'$ and $s'$, for the 88-89 and 89-90 seasons.

data strongly suggests that $\{R, D', s'\}$ are indeed independent jointly log-normal variables parametrizing the drop-size distribution, with the mean of $\log(D')$ approximately 0.4 and its standard deviation approximately 0.22, while the mean of $\log(s')$ is approximately -1 and its standard deviation approximately 0.02 . The original DSD parameters can be calculated from $\{R, D', s'\}$ using the relations

$$\mu = \frac{R^{0.04\log(R)}}{s'^2 D'^{0.7}} - 4 \tag{16}$$

$$\Lambda = \frac{R^{0.04\log(R) - 0.13}}{s'^2 D'^{1.7}} \tag{17}$$

$$N_0 = 141 \frac{\Lambda^{\mu+4.67}}{\Gamma(\mu + 4.67)} R \tag{18}$$

Additional systematic DSD measurements from other tropical locations should prove particularly useful in confirming these observations.

# 3 Conclusions

Based on two years' worth of data from Darwin, one can parametrize drop-size distributions using the variables $R$, $D'$ and $s'$ defined above, *and* assume that these parameters are independent and jointly log-normally distributed. Moreover, the variance of $D'$ is relatively small, and that of $s'$ is very small. These properties should make this parametrization particularly useful in retrieval problems.

# 4 Acknowledgements

We are very grateful to Otto Thiele and the TRMM ground truth program, and in particular to Tom Keenan, David s1101'{ and David Wolff for providing the data. This work was )(:1"-101'111('(1 at the Jet Propulsion Laboratory, California Institute Of' Technology, under contract with the National Aeronautics and Space Administration.

# 5 References

**D. Atlas and C.W. Ulbrich, 1977:** Path and area-integrated rainfall measurement by microwave attenuation in the 1-3 cm band, *J. Appl. Meteor.* **16**, pp. 1322-1331.

**J.W.F. Goddard and S.M. Cherry, 1984:** The ability Of dual-polarization radar (co- polar linear) to predict rainfall rate and microwave attenuation, *Radio Science*, **19**, pp. 201- 208."

**J. Joss and A. Waldvogel, 1967**: A raindrop spectrograph with automatic analysis, *Pure Appl. Geophys.*, **68**, pp. 240 -246.

**T. Kawanishi, 11. Takamatsu, 'J'. Kozu, K. Okamoto and 11. Kumagai, 1993**: TRMM precipitation radar, *Preprints IGARSS '93*, *])]).* 423-425.

**Kozu 'J'., and Nakamura K., 1991**: Rainfall parameter estimation from dual measurements combining reflectivity profile and path-integrated attenuation, *J. Atmos. Oceanic Technol.*, **8**, pp. 259-270.

.1. S. **Marshall ant] W.M. K. Palmer,** 1948: The distribution of rain drops with size, *J. Meteor.*, **5**, PP. 165-166.

**B.E. Sheppard, and 1'.1. Joe, 1994**: Comprai son of raindrop size distribution measurements by a Joss-Waldvogel disdrometer, a PMS 2DG spectrometer, and a 1'OSS Doppler radar, *J. Atmos. Oceanic Technol.*, **11**, *])]).* 874-887.

**C.W. Ulbrich,** 1983: Natural variations in the analytical form of the raindrop size distribution, on computer simulations of dual-measurement radar methods, *J. Climate Appl. Meteor.*, 22, PP. 1764-1775.

**C.W. Ulbrich,** 1992: Effects of drop-size distribution truncation on computer simulations of dual-measurement radar methods, *J. Appl. Meteor.*, **31**, *])]).* 689-699.

## Figure captions

Figure 1a: Simultaneous $(D^*, s^*)$ occurences during the 88-89 season, $D^*$ in millimeters.

Figure 1b: Simultaneous $(D^*, s^*)$ occurences during the 89-90 season, $D^*$ in millimeters.

Figure 1c: Simultaneous $(R, D^*)$ occurences during the 88-89 season, $R$ in mm/hr and $D^*$ in mm.

Figure 1d: Simultaneous $(R, D^*)$ occurences during the 89-90 season, $R$ in mm/hr and $D^*$ in mm.

Figure 1e: Simultaneous $(R, s^*)$ occurences during the 88-89 season, $R$ in mm/hr.

Figure 1f: Simultaneous $(R, s^*)$ occurences during the 89-90 season, $R$ in mm/hr.

Figure 2a: Simultaneous $(D', s')$ values during the 88-89 season.

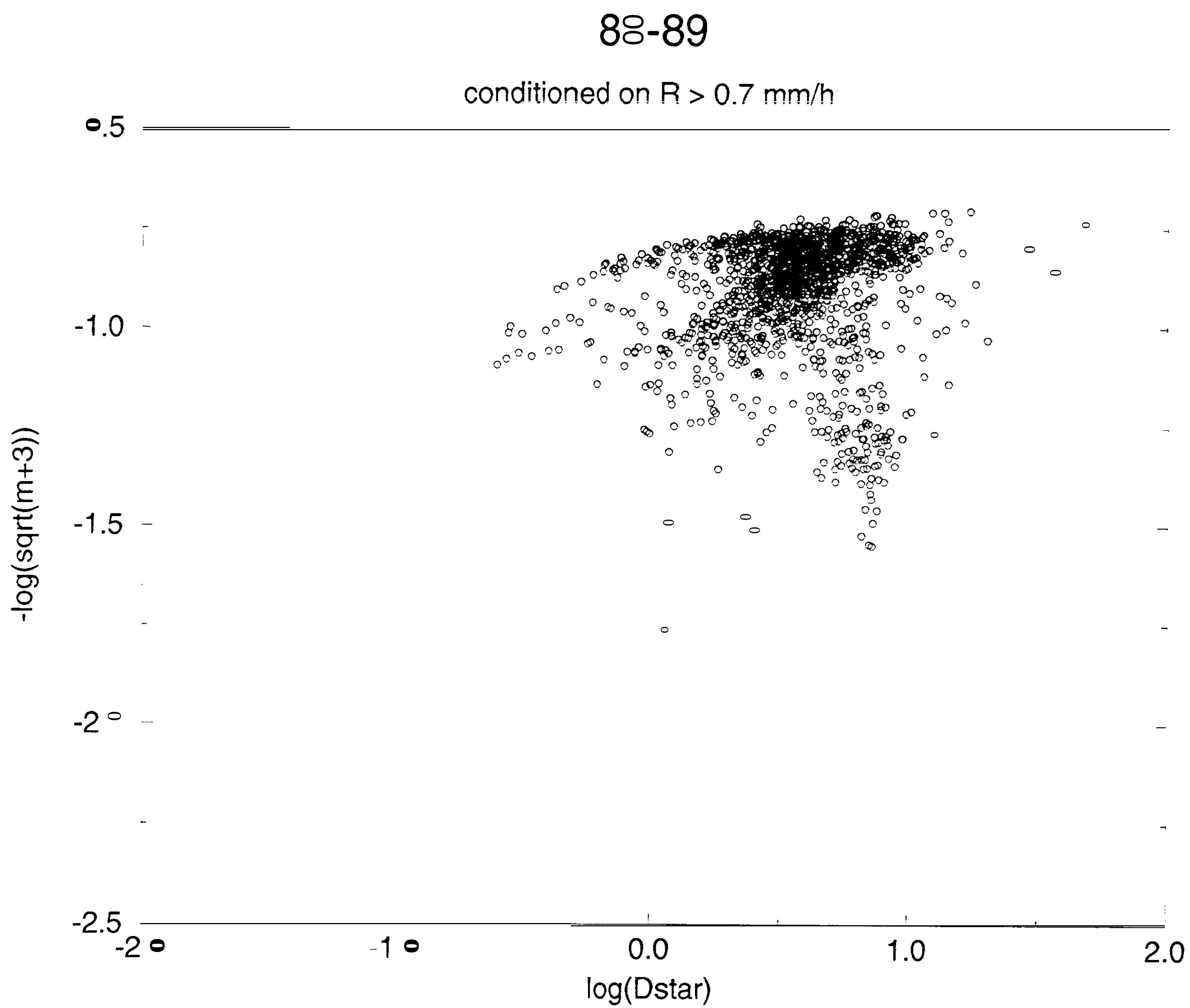
Figure 2b: Simultaneous $(D', s')$ values during the 89-90 season.

Figure 2c: Simultaneous $(R, D')$ values during the 88-89 season, $R$ in mm/hr.

Figure 2d: Simultaneous $(R, D')$ values during the 89-90 season, $R$ in mm/hr.

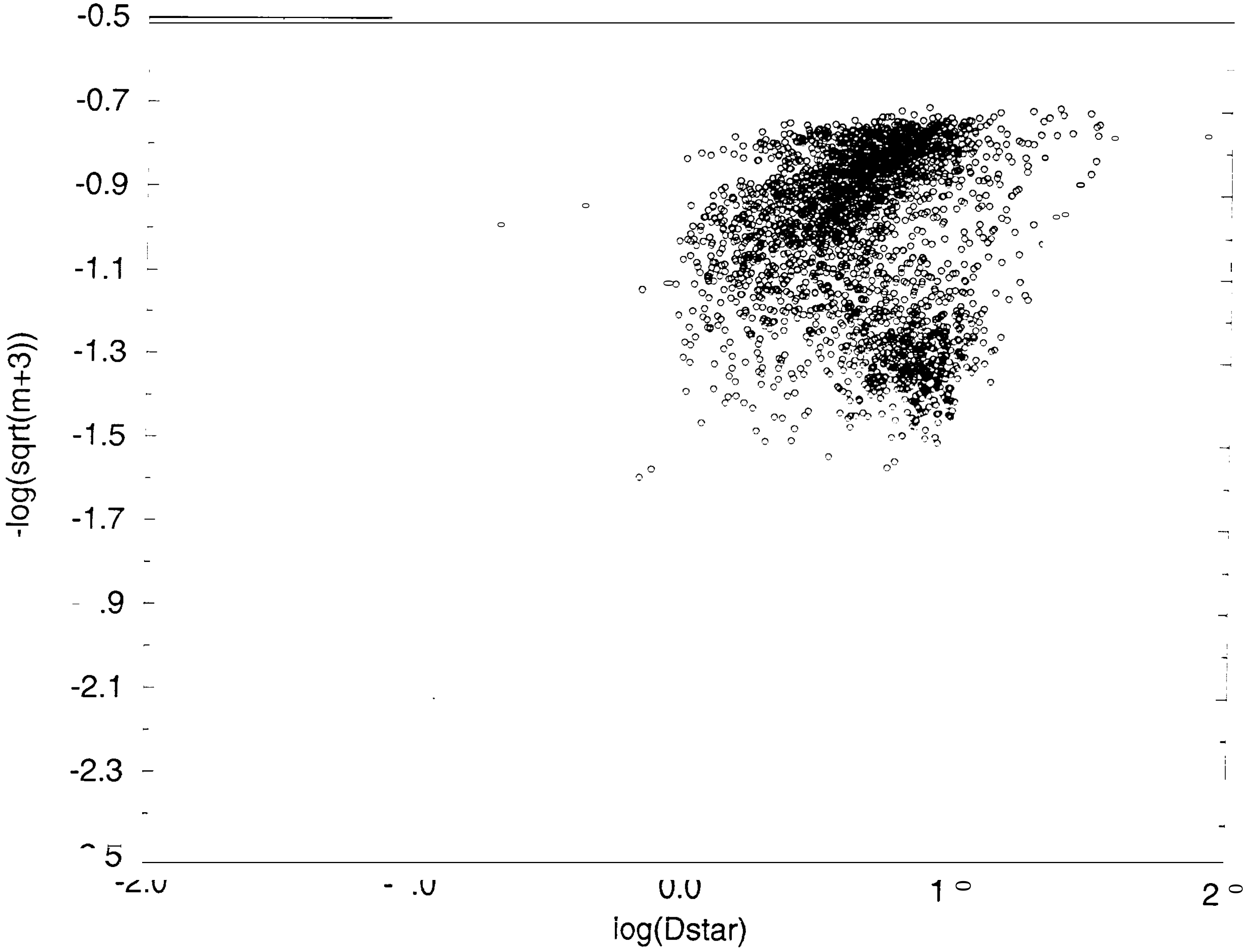
Figure 2e: Simultaneous $(R, s')$ values during the 88-89 season, $R$ in mm/hr.

Figure 2f: Simultaneous $(R, s')$ values during the 89-90 season, $R$ in mm/hr.


88-89
conditioned on R > 0.7 mm/h
-log(sqrt(m+3))
-0.5
-1.0
-1.5
-2
-2.5
-2
-1
0.0
1.0
2.0
log(Dstar)

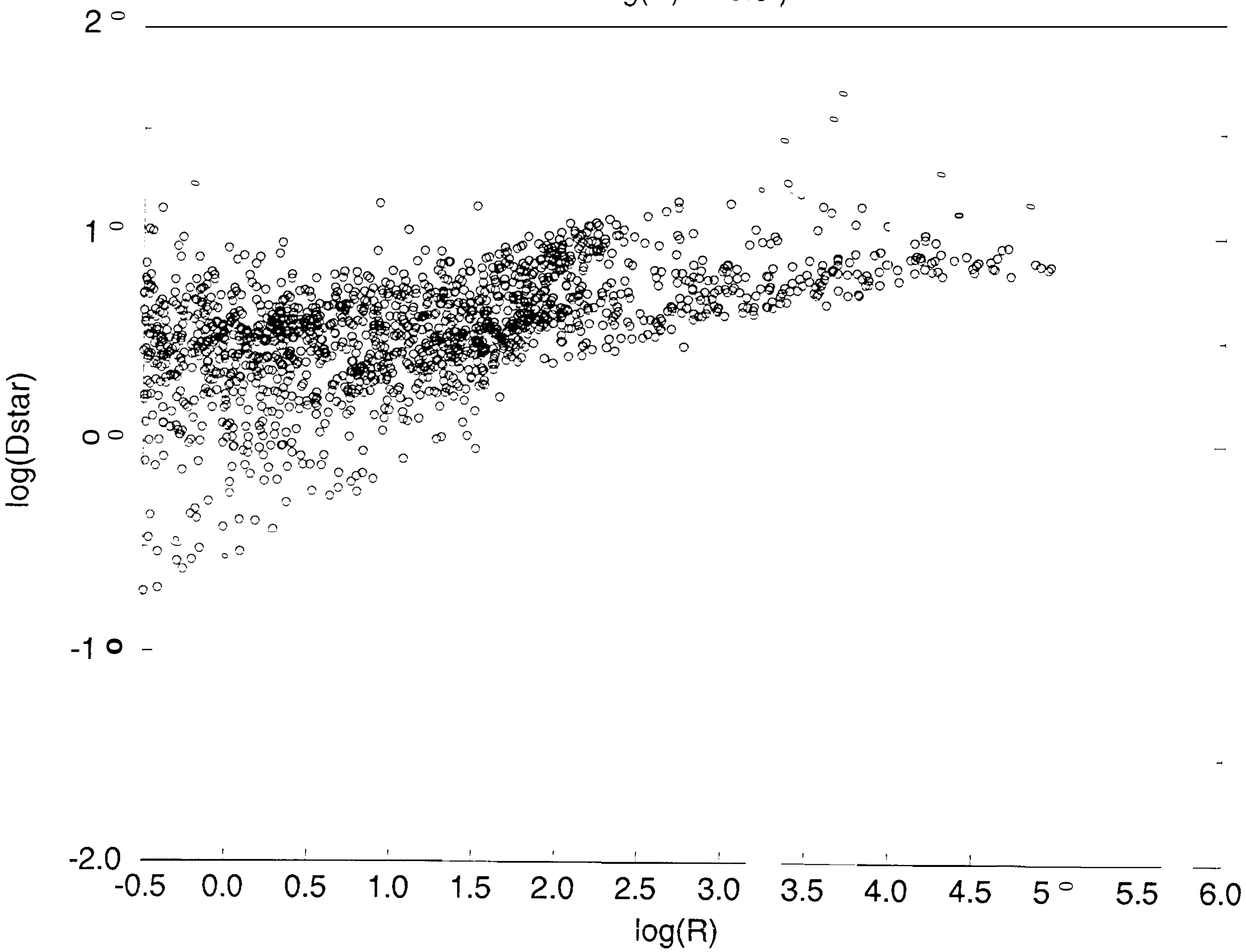
## 89-90

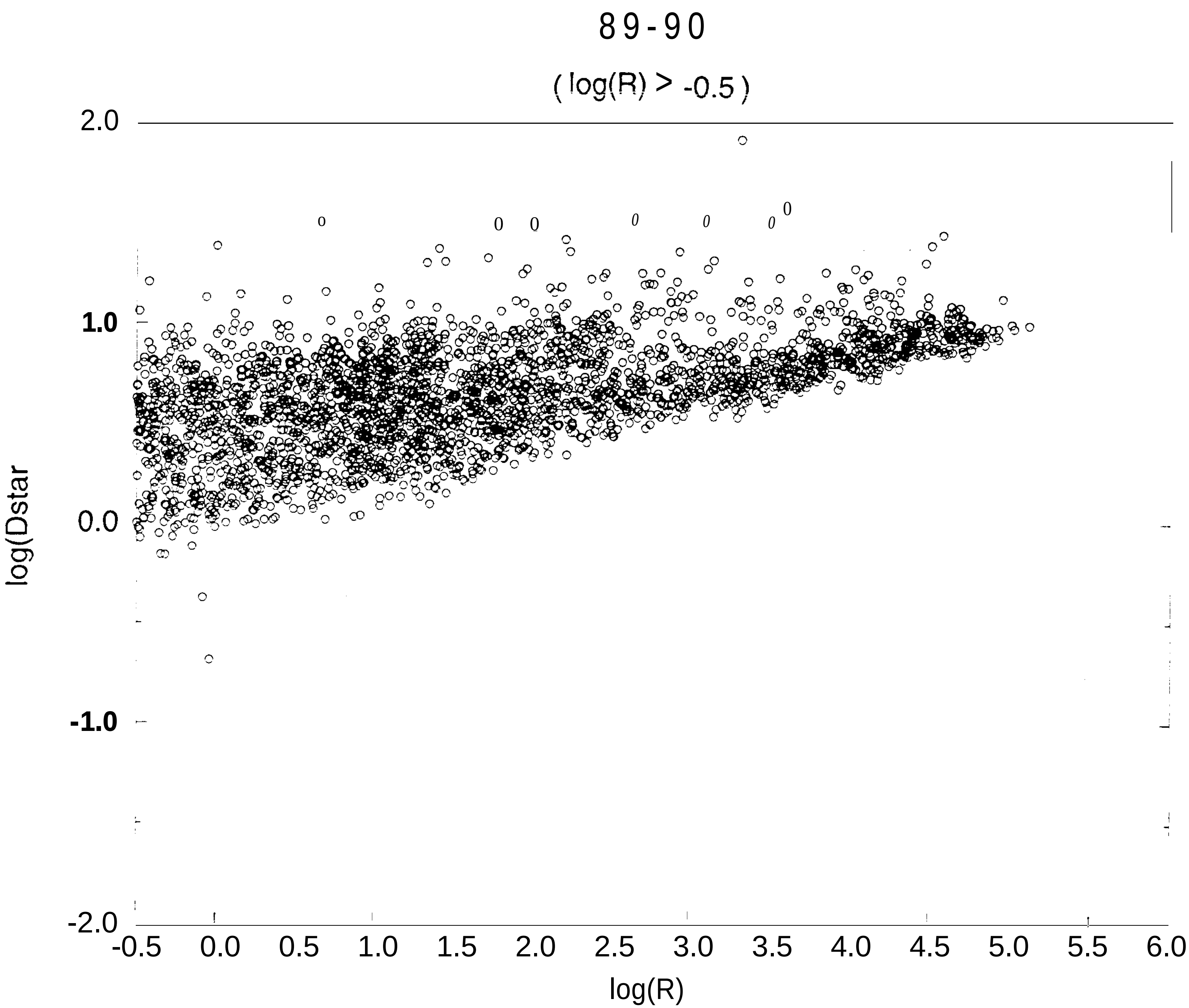
conditioned on R > 0.7 mm/hr



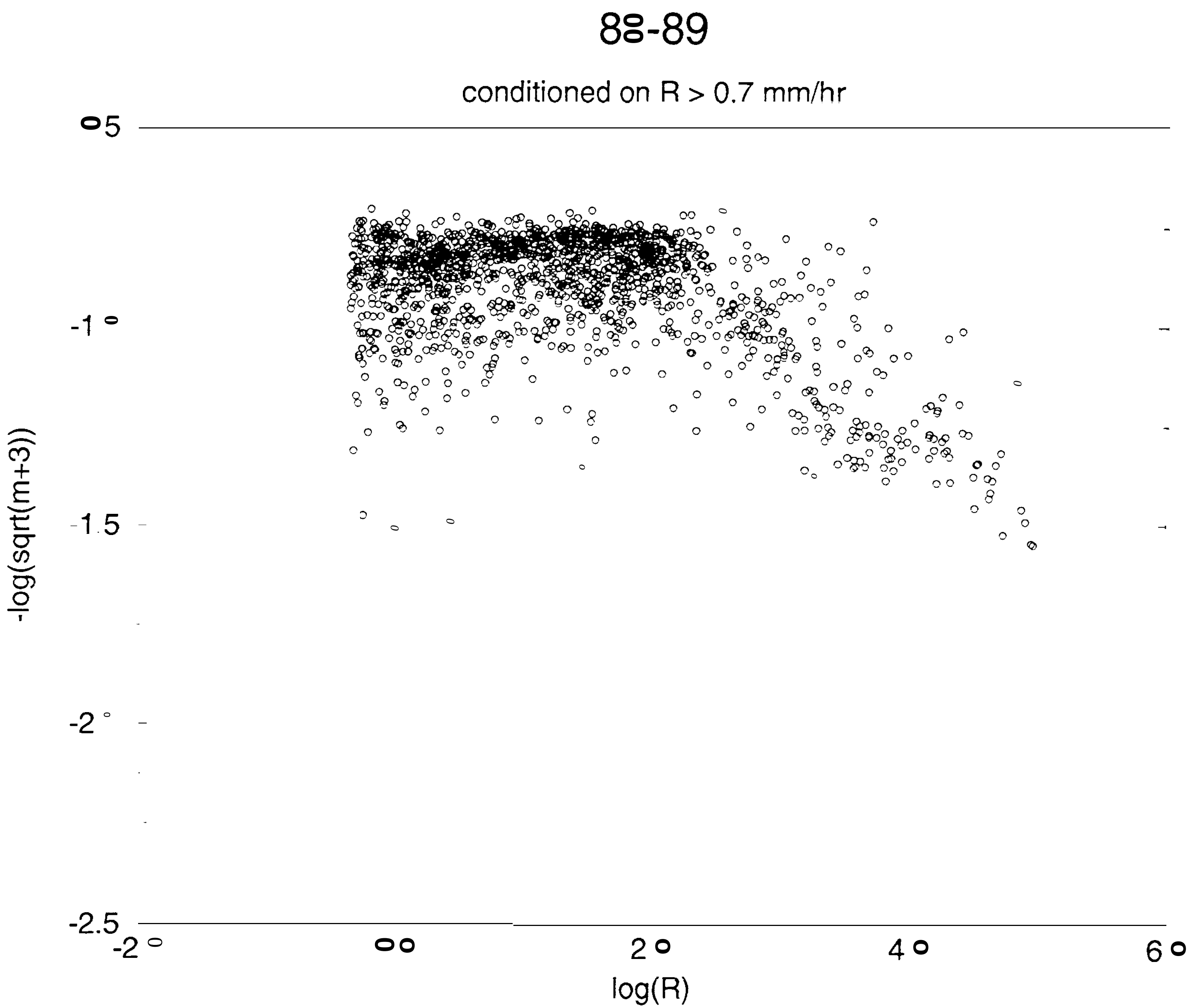
89-90

log(R) > -0.5 )



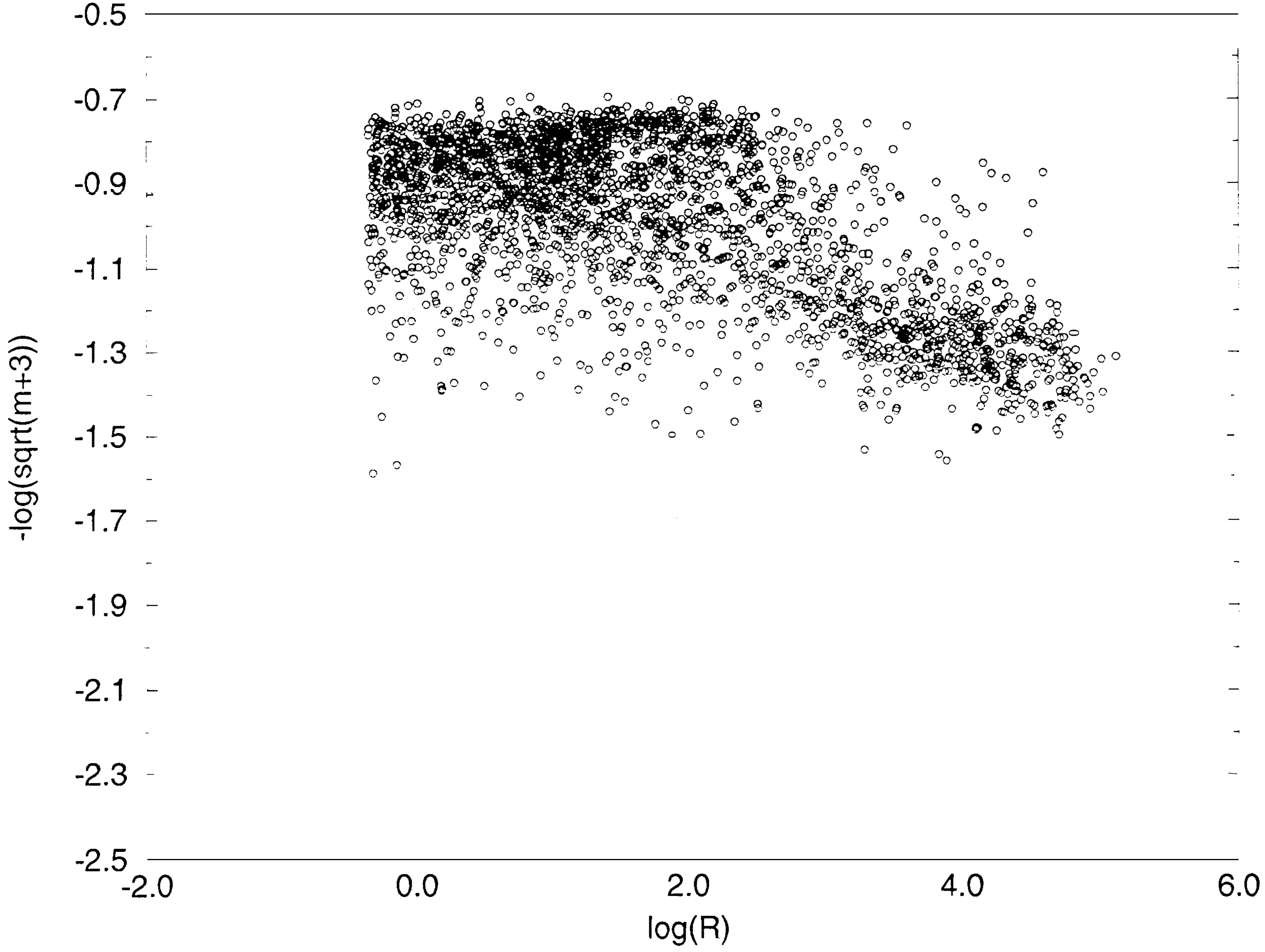


89-90
( log(R) > -0.5 )
2.0
1.0
0.0
-1.0
-2.0
log(Dstar
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
log(R)


88-89
conditioned on R > 0.7 mm/hr
-log(sqrt(m+3))
log(R)
-0.5
-1
-1.5
-2
-2.5
-2
0
2
4
6

# 89-90

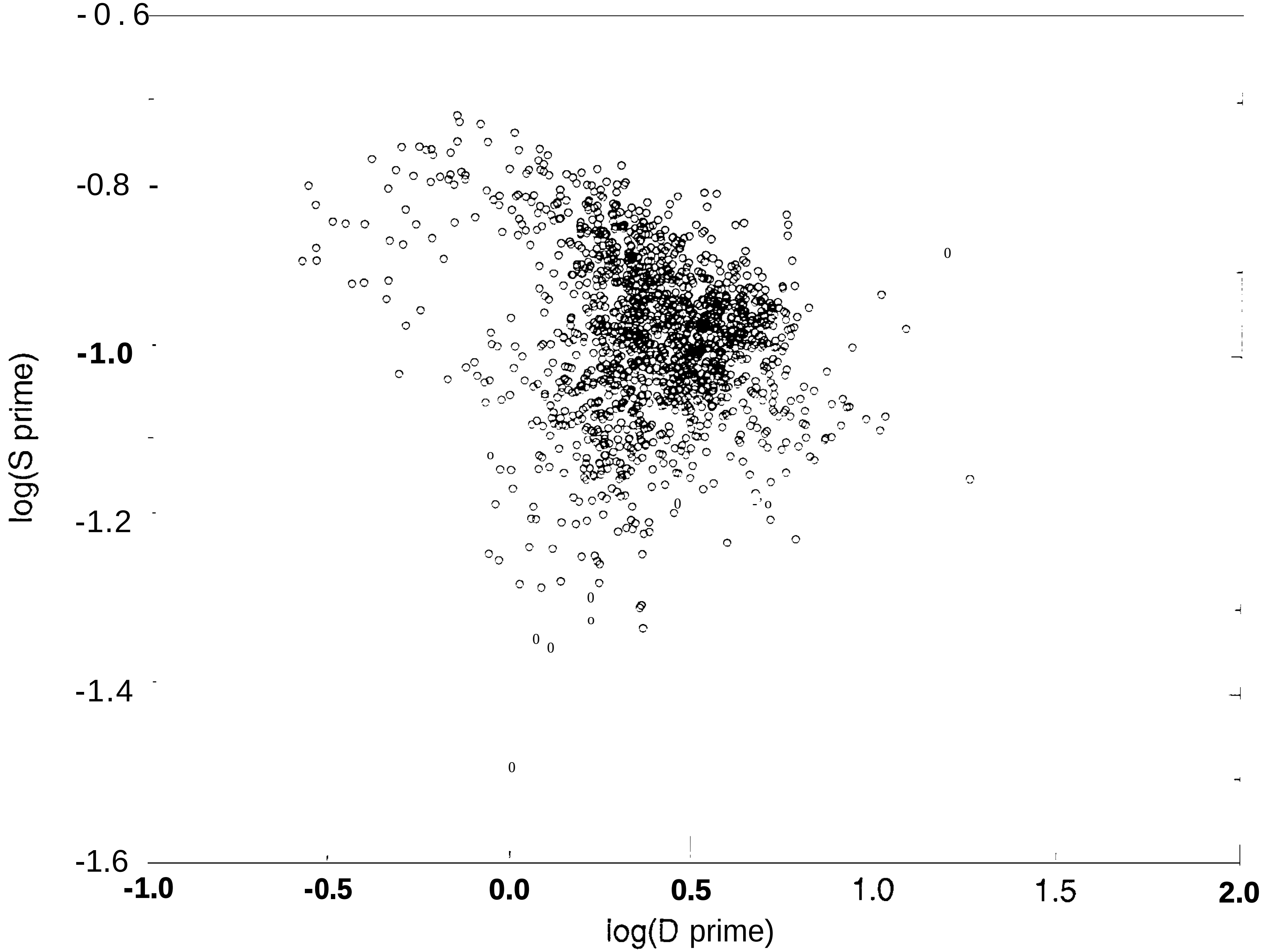
conditioned on R > 0.7 mm/hr



2a

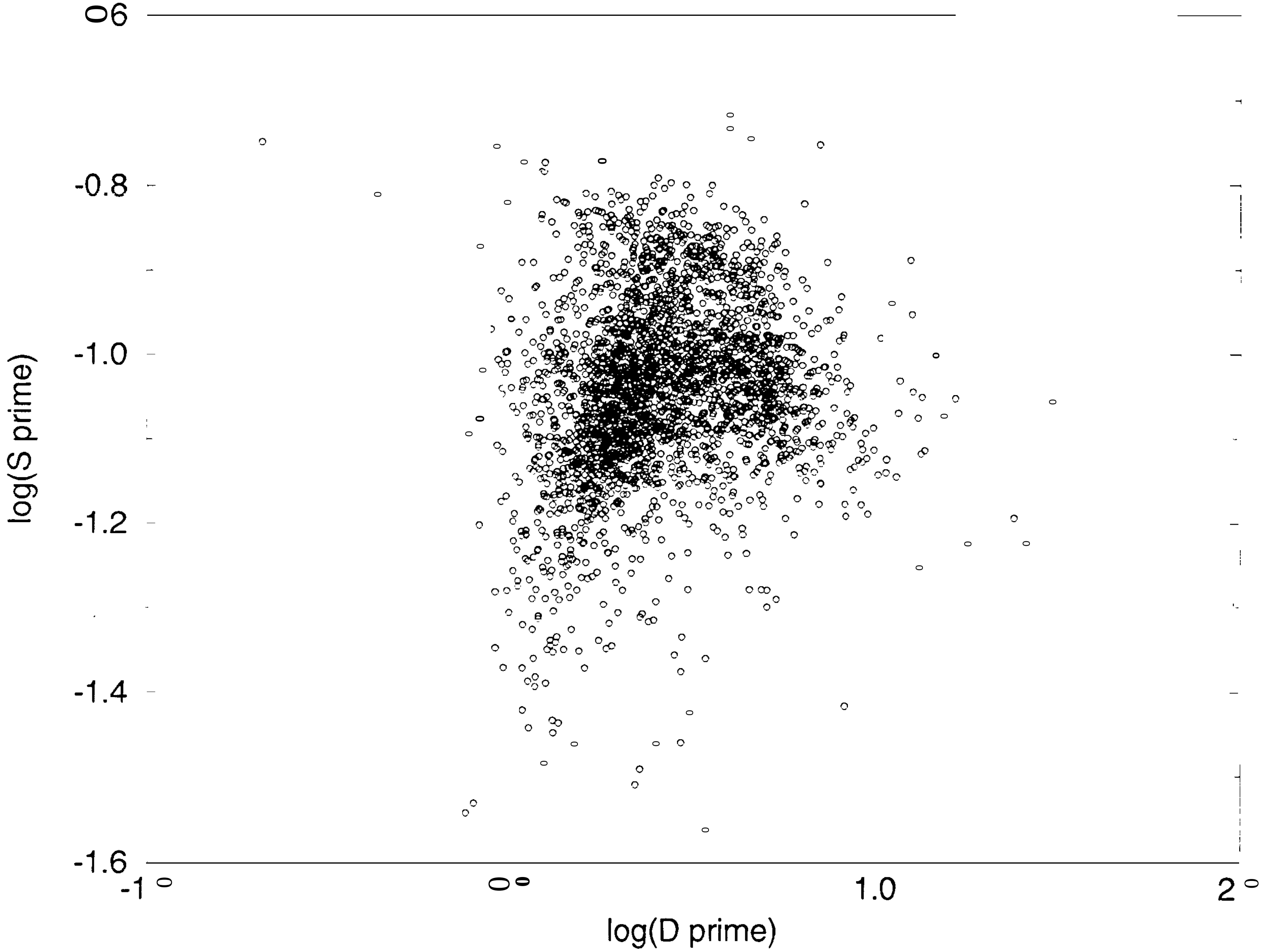
# 88-89

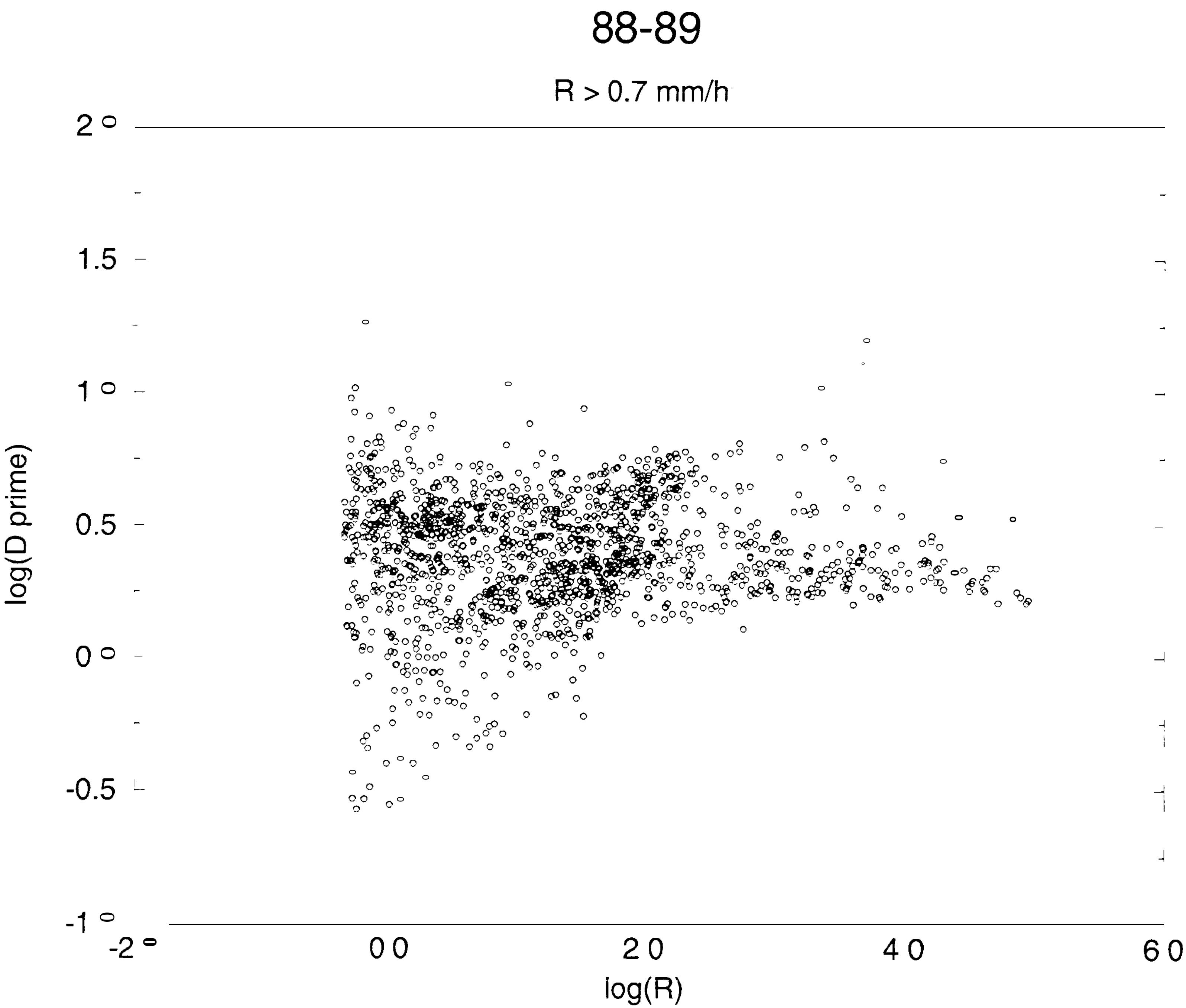
**conditioned** on R > 0.7 mm/hr



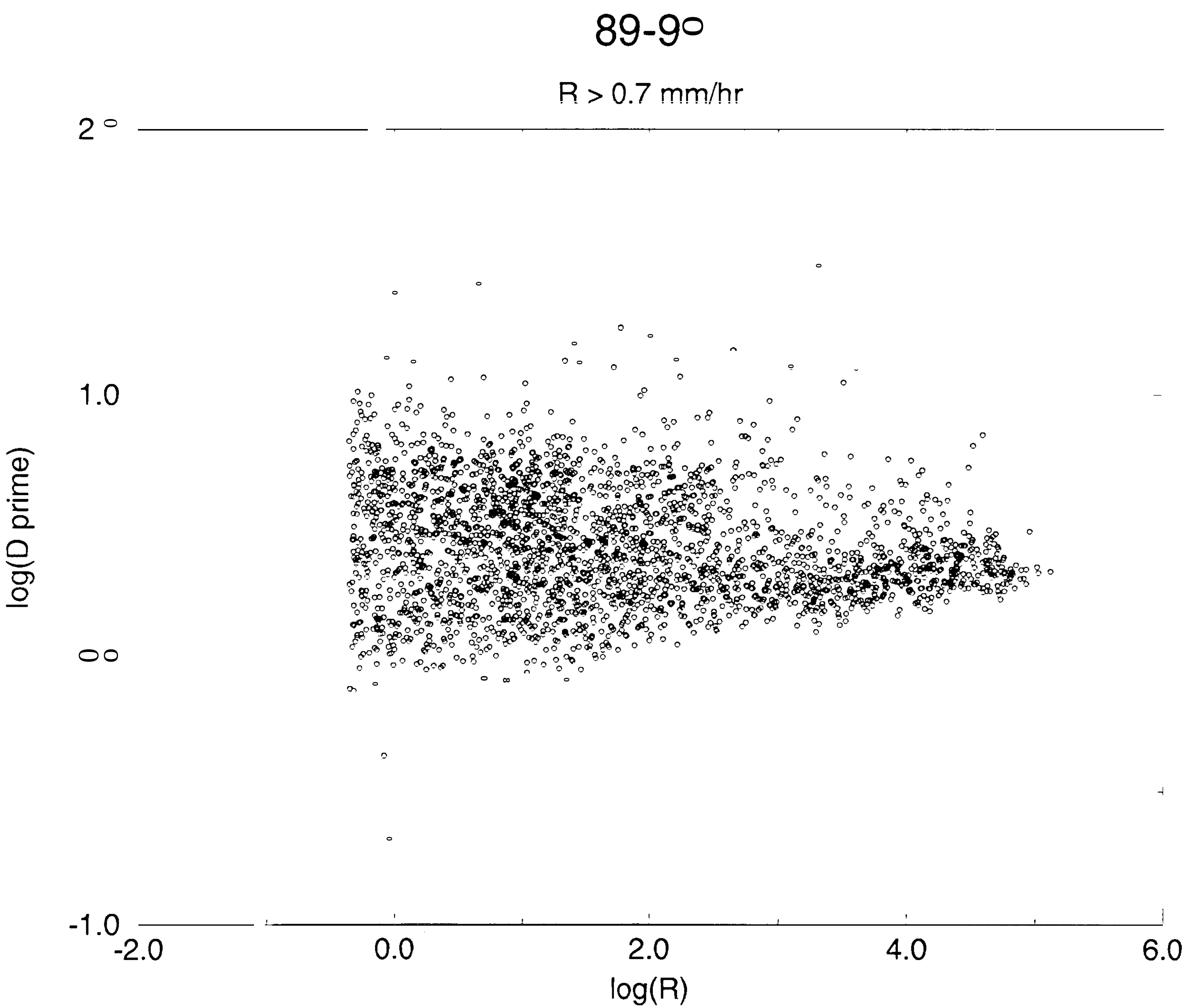
## 89-90

conditioned on R > 0.7 mm/hr



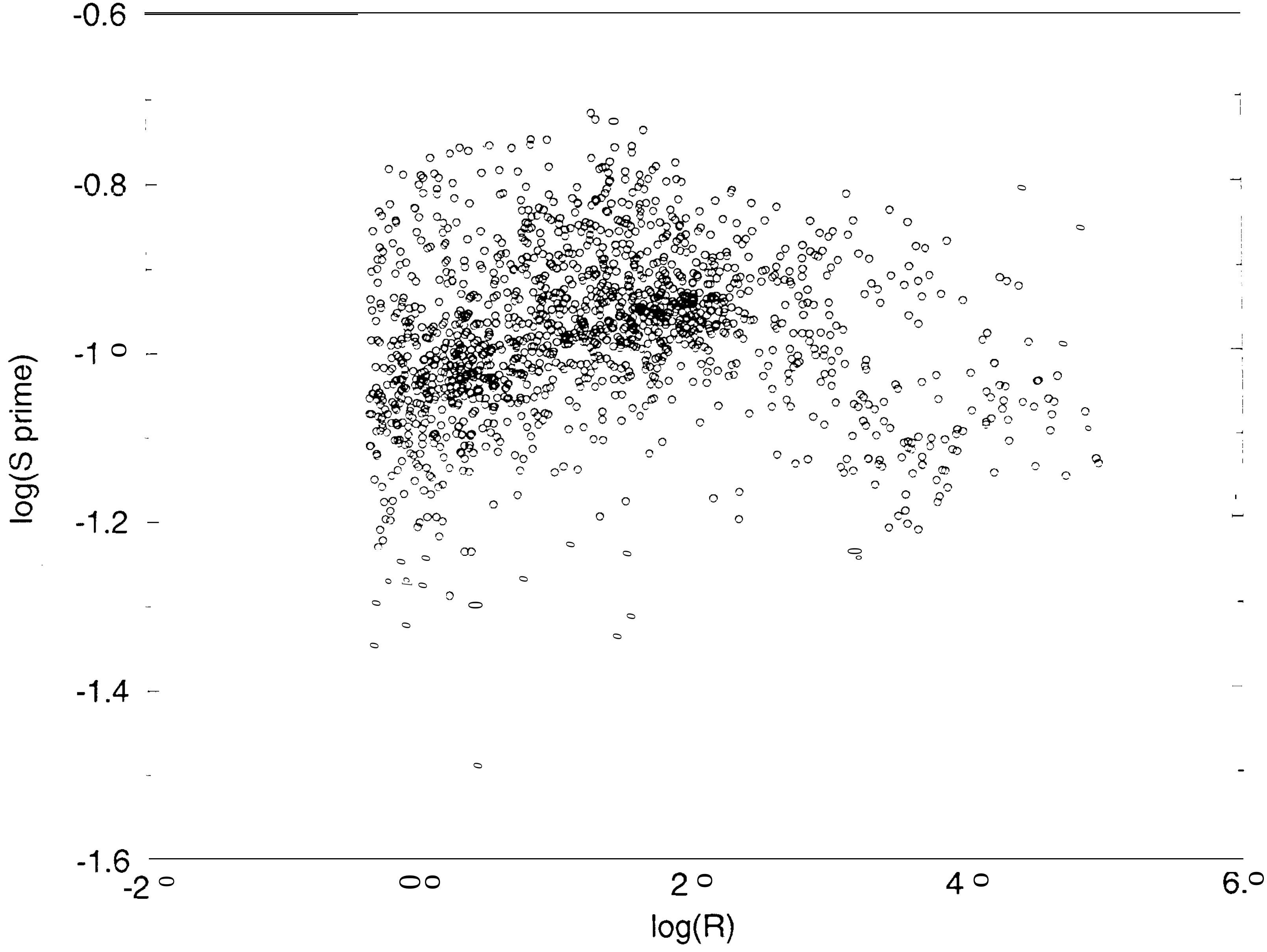


88-89
R > 0.7 mm/h
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
log(D prime)
-2.0
0.0
2.0
4.0
6.0
log(R)


89-9o
R > 0.7 mm/hr
2
1.0
0
-1.0
log(D prime)
-2.0
0.0
2.0
4.0
6.0
log(R)

88-89

R > 7 mm/hr



# 89-90

R > 0.7 mm/hr